

INVERTIBILITY OF RANDOM SUBMATRICES VIA TAIL-DECOUPLING AND A MATRIX CHERNOFF INEQUALITY

STÉPHANE CHRÉTIEN AND SÉBASTIEN DARSE

ABSTRACT. Let X be a $n \times p$ real matrix with coherence $\mu(X) = \max_{j \neq j'} |X_j^t X_{j'}|$. We present a simplified and improved study of the quasi-isometry property for most submatrices of X obtained by uniform column sampling. Our results depend on $\mu(X)$, the operator norm $\|X\|$ and the dimensions with explicit constants, which improve the previously known values by a large factor. The analysis relies on a tail-decoupling argument, of independent interest, and a recent version of the Non-Commutative Chernoff inequality (NCCI).

1. INTRODUCTION

1.1. Problem statement. Let $\mathbb{R}^{n \times p}$ denote the set of all $n \times p$ real matrices. For any $M \in \mathbb{R}^{n \times p}$, we denote by M^t its transpose and by $\|\cdot\|$ its operator norm:

$$\|M\| := \max_{x \in \mathbb{R}^p, \|x\|_2=1} \|Mx\|_2, \quad \|x\|_2^2 = x^t x.$$

Let $X \in \mathbb{R}^{n \times p}$ and T be a random index subset of size s of $\{1, \dots, p\}$ drawn from the uniform distribution. Let X_T denote the submatrix obtained by extracting the columns X_j 's of X indexed by $j \in T$. We say that X_T is an r_0 -quasi-isometry if $\|X_T^t X_T - \text{Id}\| \leq r_0$ (quasi-isometry property). The goal of this paper is to propose a new upper bound for the probability that the submatrix X_T fails to be an r_0 -quasi-isometry. In the sequel, we assume that the columns of X have unit norm.

Proving that the quasi-isometry property holds with high probability has applications in Compressed Sensing and high-dimensional statistics based on sparsity. The uniform version of the quasi-isometry property, i.e., satisfied for all possible T 's, is called the Restricted Isometry Property (RIP) and has been widely studied for independent, identically distributed (i.i.d.) sub-Gaussian matrices [7]. Recent works such as [2] proved that the quasi-isometry property holds with high probability for matrices with sufficiently small coherence $\mu(X) := \max_{j \neq j'} |X_j^t X_{j'}|$. Unlike checking the RIP, computing $\mu(X)$ can be achieved in polynomial time. Such types of result are therefore of great potential interest for a wide class of problems involving high-dimensional linear or nonlinear regression models.

Let $\{\delta_j\}$ denote a sequence of i.i.d. Bernoulli 0–1 random variables with expectation δ . Let R denote the square diagonal "selector matrix" whose j^{th} diagonal entry is δ_j . Following the landmark papers of Bourgain and Tzafriri [1] (see also [3]) and Rudelson [8], Tropp [10] established, in particular, a bound for $(\mathbb{E}\|R(X^t X - \text{Id})R\|^\rho)^{1/\rho}$, $\rho \in [2, \infty)$. As in [9], the proof heavily

relies on the Non-Commutative Khintchine inequality. Using Tropp's result, Candès and Plan proved in [2, Theorem 3.2] that X_T is a $1/2$ -quasi-isometry with probability greater than $1 - p^{-2\log(2)}$ when $s \leq p/(4\|X\|^2)$ and the coherence $\mu(X)$ is sufficiently small. The quasi-isometry property for $r_0 = \frac{1}{2}$ then holds with high probability under easily-checked assumptions on X .

1.2. Our contribution. The present paper aims at giving a more precise and self-contained version of Theorem 3.2 in [2]. Our result yields explicit constants, which improve the previously known values by a large factor. The analysis relies on a tail-decoupling argument, of independent interest, and a recent version of a Non-Commutative Chernoff inequality (NCCI) [11].

1.3. Additional notations. For $S \subset \{1, \dots, p\}$, we denote by $|S|$ the cardinality of S . Given a vector $x \in \mathbb{R}^p$, we set $x_T = (x_j)_{j \in T} \in \mathbb{R}^{|T|}$.

We denote by $\|M\|_{1 \rightarrow 2}$ the maximum l_2 -norm of a column of $M \in \mathbb{R}^{n \times p}$ and $\|M\|_{\max}$ is the maximum absolute entry of M .

In the present paper, we consider the 'hollow Gram' matrix H :

$$(1.1) \quad H = X^t X - \text{Id}.$$

In the sequel, R' will always denote an independent copy of the selector matrix R . Let R_s be a diagonal matrix whose diagonal is a random vector $\delta^{(s)}$ of length p , uniformly distributed on the set of all vectors with s components equal to 1 and $p - s$ components equal to 0. Notice that when $\delta = s/p$, the support of the diagonal of R has cardinality close to s with high probability, by a standard concentration argument.

2. PRELIMINARY RESULTS

2.1. On Rademacher chaos of order 2. Let $\{\eta_i\}$ be a sequence of i.i.d. Rademacher random variables. Theorem 3.2.2 in [6, p.113] gives the following general result: a Banach-valued homogeneous chaos X of order d

$$X = \sum_{1 \leq i_1 < \dots < i_d \leq p} X_{i_1 \dots i_d} \eta_{i_1} \dots \eta_{i_d}$$

verifies $(\mathbb{E}\|X\|^q)^{\frac{1}{q}} \leq \left(\frac{q-1}{p-1}\right)^{d/2} (\mathbb{E}\|X\|^p)^{\frac{1}{p}}$, $1 < p < q < \infty$.

We give an elementary proof in the real case with $d = 2$ and $q = 2p = 4$, which yields a better constant.

Lemma 2.1. *Let $x_{ij} \in \mathbb{R}$, $1 \leq i, j \leq p$. The homogeneous Rademacher chaos of order 2: $\xi = \sum_{i < j} x_{ij} \eta_i \eta_j$ verifies*

$$(2.2) \quad \mathbb{E} \xi^4 \leq 9 (\mathbb{E} \xi^2)^2.$$

Proof. The multinomial formula applied to ξ raised to the positive power q , gives

$$(2.3) \quad \xi^q = \sum \frac{q!}{\prod \alpha_{ij}!} \prod x_{ij}^{\alpha_{ij}} (\eta_i \eta_j)^{\alpha_{ij}},$$

where the sum is over all integers α_{ij} 's, $i < j$, such that $\sum \alpha_{ij} = q$, and the products are over all the indices (i, j) , $i < j$, ordered via the lexicographical order, still denoted by ' $<$ '. As from now, let these conventions hold.

Case $q = 2$ — The partitions of 2 are $2 + 0's$ and $1 + 1 + 0's$. Consider the partition $1 + 1 + 0's$, say $\alpha_{kl} = \alpha_{k'l'} = 1$ for some 4-uple (k, l, k', l') with $k \leq k'$. We have $(k, l) \neq (k', l')$, $k < l$ and $k' < l'$. Thus,

$$\mathbb{E}[\eta_k \eta_l \eta_{k'} \eta_{l'}] = \begin{cases} \mathbb{E}[\eta_k] \mathbb{E}[\eta_l \eta_{k'} \eta_{l'}] (= 0) & \text{if } k < k' \\ \mathbb{E}[\eta_k^2] \mathbb{E}[\eta_l] \mathbb{E}[\eta_{l'}] (= 0) & \text{if } k = k'. \end{cases}$$

Therefore, $\mathbb{E} \xi^2$ only depends on the partition $2 + 0's$, and one has

$$(2.4) \quad \mathbb{E} \xi^2 = \sum_{i < j} x_{ij}^2.$$

Case $q = 4$ — The partitions of 4 are 4 , $2 + 2$, $3 + 1$, $2 + 1 + 1$ and $1 + 1 + 1 + 1$ (we now omit the zeros).

First, using the same arguments as in the case $q = 2$, we show that the terms in $\mathbb{E} \xi^4$ corresponding to the partitions $3 + 1$ and $2 + 1 + 1$ vanish.

Second, the partitions $1 + 1 + 1 + 1$ involve four different couples (i, i') , (j, j') , (k, k') and (l, l') (recall that $i < i'$, etc., and that the couples are lexicographically ordered). The only terms corresponding to the partitions $1 + 1 + 1 + 1$ whose expectation does not vanish are of the form

$$x_{i_1 i'_1} x_{i_1 i'_2} x_{i_2 i'_1} x_{i_2 i'_2} \eta_{i_1}^2 \eta_{i'_1}^2 \eta_{i_2}^2 \eta_{i'_2}^2 = x_{i_1 i'_1} x_{i_1 i'_2} x_{i_2 i'_1} x_{i_2 i'_2},$$

i.e., the four couples $(i_1, i'_1) < (i_1, i'_2) < (i_2, i'_1) < (i_2, i'_2)$ are the vertices of a rectangle into the upper off diagonal part of the matrix (x_{ij}) . We denote by \mathcal{R} the set of all these rectangles whose vertices are lexicographically ordered.

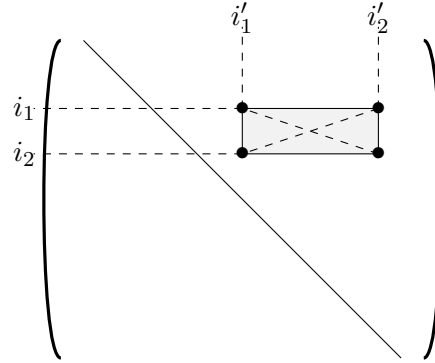


FIGURE 1. The matrix (x_{ij}) where a 'rectangle' of \mathcal{R} is drawn.

Finally, the α_{ij} 's corresponding to the partitions 4 and $2 + 2$ are even: $\alpha_{ij} = 2\beta_{ij}$, with $\sum \beta_{ij} = 2$. Therefore

$$\mathbb{E} \xi^4 = \sum \frac{4!}{\prod (2\beta_{ij})!} \prod x_{ij}^{2\beta_{ij}} + \sum_{\mathcal{R}} 4! x_{i_1 i'_1} x_{i_1 i'_2} x_{i_2 i'_1} x_{i_2 i'_2} := A + B.$$

But

$$A \leq 3 \sum \frac{2!}{\prod \beta_{ij}!} \prod (x_{ij}^2)^{\beta_{ij}} = 3 \left(\sum_{i < i'} x_{ii'}^2 \right)^2,$$

and

$$B \leq \frac{4!}{2} \sum_{\mathcal{R}} \left(x_{i_1 i'_1}^2 x_{i_2 i'_2}^2 + x_{i_1 i'_2}^2 x_{i_2 i'_1}^2 \right) \leq 6 \sum_{\substack{i < i', j < j' \\ (i, i') < (j, j')}} 2! x_{ii'}^2 x_{jj'}^2 = 6 \left(\sum_{i < i'} x_{ii'}^2 \right)^2.$$

The second inequality for B stems from relaxing the constraints induced by \mathcal{R} and illustrated in Fig. 2.1. Using (2.4), we obtain the desired result. \square

Remark 2.2. *The ratio $\mathbb{E} \xi^4 / (\mathbb{E} \xi^2)^2$ will be used in the proof of Prop. 4.1. We gain a factor 9 compared to the constant $(\frac{4-1}{2-1})^{\frac{2}{2} \cdot 4} = 81$.*

2.2. A Non-Commutative Chernoff inequality. We will also need a corollary of a Matrix Chernoff's inequality recently established in [11].

Theorem 2.3. (Matrix Chernoff Inequality [11]) *Let X_1, \dots, X_p be independent random positive semi-definite matrices taking values in $\mathbb{R}^{d \times d}$. Set $S_p = \sum_{j=1}^p X_j$. Assume that for all $j \in \{1, \dots, p\}$, $\|X_j\| \leq B$ a.s. and*

$$\|\mathbb{E} S_p\| \leq \mu_{\max}.$$

Then, for all $r \geq e \mu_{\max}$,

$$\mathbb{P}(\|S_p\| \geq r) \leq d \left(\frac{e \mu_{\max}}{r} \right)^{r/B}.$$

(Set $r = (1 + \delta)\mu_{\max}$ and use $e^\delta \leq e^{1+\delta}$ in Theorem 1.1 [11].)

3. MAIN RESULTS

3.1. Singular-value concentration theorem.

Theorem 3.1. *Let $r \in (0, 1)$, $\alpha \geq 1$. Let us be given a full-rank matrix $X \in \mathbb{R}^{n \times p}$ and a positive integer s , such that*

$$(3.5) \quad \mu(X) \leq \frac{r}{2(1 + \alpha) \log p}$$

$$(3.6) \quad s \leq \frac{r^2}{4(1 + \alpha)e^2} \frac{p}{\|X\|^2 \log p}.$$

Let $T \subset \{1, \dots, p\}$ be a set with cardinality s , chosen randomly from the uniform distribution. Then the following bound holds:

$$(3.7) \quad \mathbb{P}(\|X_T^t X_T - \text{Id}_s\| \geq r) \leq \frac{216}{p^\alpha}.$$

3.2. Remarks on the various constants.

The constant 216 stems from the following decomposition: 2 (poissonization) \times 36 (decoupling) \times 3 (union bound). This constant might look large. However, in many statistical applications as in sparse models, p is often assumed to be very large.

Let us now compare the constants C_s and C_μ in the inequalities

$$(3.8) \quad \mu(X) \leq \frac{C_\mu}{\log p}$$

$$(3.9) \quad s \leq C_s \frac{p}{\|X\|^2 \log p},$$

to the one of [2]. The larger C_s and C_μ are, the better the result is.

One of the various constraints on the rate α in [2] is given by the theorem of Tropp in [10]. In this setting, $\alpha = 2 \log 2$ and $r_0 = 1/2$, the author's choice of $1/2$ being unessential. To obtain such a rate α , they need to impose the r.h.s. of (3.15) in [2] to be less than $1/4$, that is $30C_\mu + 13\sqrt{2C_s} \leq \frac{1}{4}$. This yields $C_s < 1.19 \times 10^{-4}$. Choosing C_s close to 1.19×10^{-4} , e.g. $C_s \simeq 1.18 \cdot 10^{-4}$, we obtain:

$$C_s \simeq 1.18 \cdot 10^{-4}, \quad C_\mu \simeq 1.7 \cdot 10^{-3}.$$

Our theorem allows to choose any rate $\alpha > 0$. To make a fair comparison, let us choose $\alpha = 2 \log 2$ and $r = 1/2$. We obtain:

$$C_s \simeq 3.5 \cdot 10^{-3}, \quad C_\mu \simeq 0.1.$$

4. PROOF OF THEOREM 3.1

In order to study the invertibility condition, we want to obtain bounds for the distribution tail of random sub-matrices of $H = X^t X - \text{Id}$.

Let R' be an independent copy of R . Let us recall two basic estimates:

$$\|H\|_{1 \rightarrow 2}^2 \leq \|X\|^2, \quad \|H\|^2 \leq \|X\|^4.$$

As a preliminary, let us notice that

$$(4.10) \quad \mathbb{P}(\|R_s H R_s\| \geq r) \leq 2 \mathbb{P}(\|R H R\| \geq r),$$

which can be actually proven using the same kind of 'Poissonization argument' as in Claim (3.29) p. 2173 in [2].

To study the tail-distribution of $\|R H R\|$, we use a decoupling technique which consists of replacing $\|R H R\|$ with $\|R H R'\|$.

Proposition 4.1. *The operator norm of $R H R$ satisfies*

$$(4.11) \quad \mathbb{P}(\|R H R\| \geq r) \leq 36 \mathbb{P}(\|R H R'\| \geq r/2).$$

The main feature of this inequality is that the numerical constants are improved by a great factor when compared to the general result [5, Theorem 1 p.224] (cf. Remark 5.1). In addition to this decoupling argument, we need the following technical concentration result.

Proposition 4.2. *Let $X \in \mathbb{R}^{n \times p}$ be a full-rank matrix. For all parameters s, r, u, v such that $\frac{p r^2}{s e} \geq u^2 \geq \frac{s}{p} \|X\|^4$ and $v^2 \geq \frac{s}{p} \|X\|^2$, the following bound holds:*

$$(4.12) \quad \mathbb{P}(\|R H R'\| \geq r) \leq 3 p \mathcal{V}(s, [r, u, v]),$$

with

$$\mathcal{V}(s, [r, u, v]) = \left(e \frac{s u^2}{p r^2} \right)^{\frac{r^2}{v^2}} + \left(e \frac{s \|X\|^4}{p u^2} \right)^{u^2 / \|X\|^2} + \left(e \frac{s \|X\|^2}{p v^2} \right)^{v^2 / \mu(X)^2}.$$

We now have to analyze carefully the various quantities in Proposition 4.2 in order to obtain for $P(\|R H R'\| \geq r/2)$ a bound of the order $e^{-\alpha \log p}$.

Set $\alpha' = \alpha + 1$ and $r' = r/2$. We tune the parameters so that

$$(4.13) \quad \frac{u^2}{\|X\|^2} = \alpha' \log p$$

$$(4.14) \quad \frac{v^2}{\mu(X)^2} = \alpha' \log p$$

$$(4.15) \quad \frac{r'^2}{v^2} \geq \alpha' \log p,$$

and

$$(4.16) \quad e \frac{s}{p} \frac{\|X\|^4}{u^2} \leq e^{-1}$$

$$(4.17) \quad e \frac{s}{p} \frac{\|X\|^2}{v^2} \leq e^{-1}$$

$$(4.18) \quad e \frac{s}{p} \frac{u^2}{r'^2} \leq e^{-1}.$$

A crucial quantity turns out to be $\frac{s}{p} \|X\|^2$. Keeping in mind that the hypothesis on the coherence reads

$$(4.19) \quad \mu(X) \leq \frac{C_\mu}{\log p},$$

it is necessary to impose that s satisfies

$$(4.20) \quad \frac{s}{p} \|X\|^2 = \frac{C_s}{\log p},$$

The constants C_μ and C_s will be tuned according to several constraints. The equalities (4.13-4.14) determine the values of u and v . It remains to show that the previous inequalities are satisfied for a suitable choice of C_μ and C_s .

First, substituting (4.13) into (4.18), we obtain:

$$\alpha' \frac{s}{p} \|X\|^2 \log p \leq e^{-2} r'^2.$$

Using (4.20), it follows that

$$(4.21) \quad C_s \leq \frac{r'^2}{\alpha' e^2}.$$

Now, the bound (4.16) is satisfied if

$$\frac{e^2 C_s}{\log p} \leq \alpha' \log p.$$

Based on (4.21), it suffices to have $\frac{r'^2}{\alpha'^2} \leq \log^2 p$, that is $p \geq e > e^{r'/\alpha'}$.

Second, substituting (4.14) into (4.17), we obtain:

$$e^2 \frac{s}{p} \|X\|^2 \leq \alpha' \mu(X)^2 \log p.$$

Using (4.19) and (4.20), it follows that

$$e \sqrt{\frac{C_s}{\alpha'}} \leq C_\mu.$$

Finally, (4.14-4.15) yields $r'^2 \geq \alpha'^2 \mu(X)^2 \log^2 p$. In view of (4.19), it thus suffices to have $r' \geq \alpha' C_\mu$.

To reach the desired conclusion, in order to ensure the six previous constraints, it suffices to choose C_s and C_μ such that:

$$C_\mu \leq \frac{r'}{1+\alpha} \quad \text{and} \quad C_s \leq \min \left(\frac{r'^2}{(1+\alpha)e^2}, (1+\alpha)\frac{C_\mu^2}{e^2} \right).$$

This completes the proof of Theorem 3.1.

5. PROOF OF THE TAIL-DECOUPLING AND THE CONCENTRATION RESULT

5.1. Proof of Proposition 4.1. Let us write

$$RHR = \sum_{i \neq j} \delta_i \delta_j H_{ij}.$$

Let $\{\eta_i\}$ be a sequence of i.i.d. independent Rademacher random variables, mutually independent of $\mathcal{D} := \{\delta_i, 1 \leq i \leq p\}$. Following Bourgain and Tzafriri [1], and de la Peña and Giné [6], we construct an auxiliary random variable:

$$Z = Z(\eta, \delta) := \sum_{i \neq j} (1 - \eta_i \eta_j) \delta_i \delta_j H_{ij}.$$

Setting $Y = \sum_{i \neq j} \delta_i \delta_j H_{ij} \eta_i \eta_j$, we can write

$$(5.22) \quad Z = RHR + Y.$$

For the sake of completeness, we recall basic arguments from Corollary 3.3.8 p.12 in de la Peña and Giné [6] (applied to (5.22)) to obtain a lower bound for $\mathbb{P}(\|Z\| \geq \|RHR\|)$. (We henceforth work conditionally on \mathcal{D} .)

Hahn-Banach's theorem gives a linear form x^* on $\mathbb{R}^{p \times p}$ such that

$$(5.23) \quad \begin{aligned} \mathbb{P}(\|Z\| \geq \|RHR\| \mid \mathcal{D}) &\geq \mathbb{P}(x^*(Z) \geq x^*(RHR) \mid \mathcal{D}) \\ &\geq \mathbb{P}(x^*(Y) \geq 0 \mid \mathcal{D}). \end{aligned}$$

For any centered real random variable ξ , one obtains using Hölder's inequality twice (first with $\mathbb{E}|\xi| = 2\mathbb{E} \xi \mathbf{1}_{\xi > 0}$, second with $\mathbb{E} \xi^2 = \mathbb{E} \xi^{2/3} \xi^{4/3}$):

$$(5.24) \quad \mathbb{P}(\xi \geq 0) \geq \frac{1}{4} \frac{(\mathbb{E}|\xi|)^2}{\mathbb{E} \xi^2} \geq \frac{1}{4} \frac{(\mathbb{E} \xi^2)^2}{\mathbb{E} \xi^4}.$$

Noticing that $x^*(Y)$ is a centered homogeneous real chaos of order 2, we deduce from (5.23), (5.24) and Lemma 2.1,

$$(5.25) \quad \mathbb{P}(\|Z\| \geq \|RHR\| \mid \mathcal{D}) \geq \frac{1}{4 \times 9} = \frac{1}{36}.$$

Multiplying both sides by $\mathbf{1}_{\{\|RHR\| \geq r\}}$ and taking the expectation, one has

$$(5.26) \quad \frac{1}{36} \mathbb{P}(\|RHR\| \geq r) \leq \mathbb{P}(\|Z\| \geq r).$$

As from now, we can use similar arguments to [10, Prop. 2.1]. There is a $\eta^* \in \{-1, 1\}^p$ for which

$$\mathbb{P}(\|Z\| \geq r) = \mathbb{E} \mathbb{E} [\mathbf{1}_{\{\|Z\| \geq r\}} \mid (\eta_i)] \leq \mathbb{E} \mathbf{1}_{\{\|Z(\eta^*, \delta)\| \geq r\}} = \mathbb{P}(\|Z(\eta^*, \delta)\| \geq r).$$

Hence, setting $T = \{i, \eta_i^* = 1\}$, we can write

$$Z(\eta^*, \delta) = 2 \sum_{j \in T, k \in T^c} \delta_j \delta_k H_{jk} + 2 \sum_{j \in T^c, k \in T} \delta_j \delta_k H_{jk}.$$

Since H is hermitian, we have

$$\left\| \sum_{j \in T, k \in T^c} \delta_j \delta_k H_{jk} + \sum_{j \in T^c, k \in T} \delta_j \delta_k H_{jk} \right\| = \left\| \sum_{j \in T, k \in T^c} \delta_j \delta_k H_{jk} \right\|.$$

Now, let (δ'_i) be an independent copy of (δ_i) . Set $\tilde{\delta}_i = \delta_i$ if $i \in T$ and $\tilde{\delta}_i = \delta'_i$ if $i \in T^c$. Since the vectors (δ_i) and $(\tilde{\delta}_i)$ have the same law, we then obtain:

$$\mathbb{P}(\|Z\| \geq r) \leq \mathbb{P}\left(2 \left\| \sum_{j \in T, k \in T^c} \delta_j \delta'_k H_{jk} \right\| \geq r\right).$$

Re-introducing the missing entries in H yields

$$\mathbb{P}(\|Z\| \geq r) \leq \mathbb{P}(\|RHR'\| \geq r/2),$$

which concludes the proof of the lemma due to (5.26).

Remark 5.1. *The previous result can be seen as a special case of Theorem 1 p.224 of the seminal paper [5]. Tracing the various constants involved in this theorem, we obtained the inequality*

$$(5.27) \quad \mathbb{P}(\|RHR\| \geq r) \leq 10^3 \mathbb{P}\left(\|RHR'\| \geq \frac{r}{18}\right).$$

5.2. Proof of Proposition 4.2. We first apply the NCCI to $\|RHR'\|$ by conditioning on R .

Lemma 5.2. *The following bound holds:*

$$(5.28) \quad \begin{aligned} P(\|RHR'\| \geq r) &\leq \mathbb{P}(\|RH\| \geq u) + \mathbb{P}(\|RH\|_{1 \rightarrow 2} \geq v) \\ &\quad + p \left(e \frac{s}{p} \frac{u^2}{r^2} \right)^{\frac{r^2}{v^2}}. \end{aligned}$$

Proof. We have $\|RHR'\|^2 = \|RHR'^2HR\|$. But $R'^2 = R'$, so

$$(5.29) \quad \mathbb{P}(\|RHR'\| \geq r) = P(\|RHR'HR\| \geq r^2).$$

We will first compute the conditional probability

$$(5.30) \quad \mathbb{P}(\|RHR'HR\| \geq r^2 \mid R) := \mathbb{E}[\mathbf{1}_{\{\|RHR'HR\| \geq r^2\}} \mid R].$$

Let Z_j be the j^{th} column of RH , $j \in \{1, \dots, p\}$. Notice that

$$RHR'HR = \sum_{j=1}^p \delta'_j Z_j Z_j^t := \sum_{j=1}^p A_j.$$

Since $\sum_{j=1}^p Z_j Z_j^t = RH^2R$ and $\|Z_j Z_j^t\| = \|Z_j\|_2^2$, we then obtain

$$(5.31) \quad \|A_j\| \leq \|RH\|_{1 \rightarrow 2}^2$$

$$(5.32) \quad \left\| \sum_{j=1}^p \mathbb{E} A_j \right\| \leq \frac{s}{p} \|RH\|^2.$$

The NCCI then yields

$$(5.33) \quad \mathbb{P}(\|RHR'HR\| \geq r^2 \mid R) \leq p \left(e \frac{s}{p} \frac{\|RH\|^2}{r^2} \right)^{r^2 / \|RH\|_{1 \rightarrow 2}^2},$$

provided that

$$(5.34) \quad e \frac{s \|RH\|^2}{p r^2} \leq 1.$$

Let us now introduce the events

$$\mathcal{A} = \{\|RHR'HR\| \geq r^2\}; \quad \mathcal{B} = \{\|RH\| \geq u\}; \quad \mathcal{C} = \{\|RH\|_{1 \rightarrow 2} \geq v\}.$$

We have

$$\begin{aligned} \mathbb{P}(\mathcal{A}) &= \mathbb{P}(\mathcal{A} \mid \mathcal{B} \cup \mathcal{C}) \mathbb{P}(\mathcal{B} \cup \mathcal{C}) + \mathbb{P}(\mathcal{A} \cap \mathcal{B}^c \cap \mathcal{C}^c) \\ &\leq \mathbb{P}(\mathcal{B}) + \mathbb{P}(\mathcal{C}) + \mathbb{P}(\mathcal{A} \cap \mathcal{B}^c \cap \mathcal{C}^c). \end{aligned}$$

The identity $\mathbb{P}(\mathcal{A} \cap \mathcal{B}^c \cap \mathcal{C}^c) = \mathbb{E}[\mathbf{1}_{\mathcal{A} \cap \mathcal{B}^c \cap \mathcal{C}^c}] = \mathbb{E}[\mathbb{P}(\mathcal{A} \mid R) \mathbf{1}_{\mathcal{B}^c \cap \mathcal{C}^c}]$ concludes the lemma. \square

We now have to control the norm of $\frac{s}{p}RH^2R$, the norm of RH and the column norm of RH . Let us begin with $\|RH\| = \|HR\|$.

Lemma 5.3. *The following bounds hold:*

$$\begin{aligned} \mathbb{P}(\|HR\| > u) &\leq p \left(e \frac{s \|X\|^4}{p u^2} \right)^{u^2/\|X\|^2} \\ \mathbb{P}(\|RH\|_{1 \rightarrow 2} \geq v) &\leq p \left(e \frac{s \|X\|^2}{p v^2} \right)^{v^2/\mu(X)^2}, \end{aligned}$$

provided that $e \frac{s \|X\|^4}{p u^2}$ and $e \frac{s \|X\|^2}{p v^2}$ are less than 1.

Proof. The steps are of course the same as what we have just done in the proof of Lemma 4.1. Notice that

$$\mathbb{P}(\|RH\| > u) = \mathbb{P}(\|HR\|^2 > u^2) = \mathbb{P}(\|HRH\| > u^2).$$

The j^{th} column of H is $H_j = X^t X_j - e_j$. Moreover,

$$(5.35) \quad HRH = \sum_{j=1}^p \delta_j H_j H_j^t.$$

We have $\|H_j H_j^t\| = \|H_j\|_2^2 \leq \|H\|_{1 \rightarrow 2}^2 \leq \|X\|^2$, and

$$(5.36) \quad \left\| \sum_{j=1}^p \mathbb{E}[\delta_j H_j H_j^t] \right\| \leq \frac{s}{p} \|H\|^2 \leq \frac{s}{p} \|X\|^4.$$

We finally deduce from the NCCI that

$$(5.37) \quad \mathbb{P}(\|HRH\| \geq u^2) \leq p \left(e \frac{s \|X\|^4}{p u^2} \right)^{u^2/\|X\|^2}.$$

Let us now control the supremum ℓ_2 -norm of the columns of RH . Set

$$(5.38) \quad M = \sum_{k=1}^p \delta_k \text{diag}(H_k H_k^t).$$

Notice that

$$\|RH\|_{1 \rightarrow 2}^2 = \max_{k=1}^p \|(RH)_k\|_2^2 = \|\text{diag}((RH)^t RH)\| = \|\text{diag}(H^t RH)\|.$$

Thus,

$$\|RH\|_{1 \rightarrow 2}^2 = \left\| \text{diag} \left(\sum_{k=1}^p \delta_k (H^t)_k H_k^t \right) \right\|.$$

Using symmetry of H and interchanging the summation and the "diag" operation, we obtain that $\|RH\|_{1 \rightarrow 2}^2 = \|M\|$. Moreover, we have for all $k \in \{1, \dots, p\}$,

$$(5.39) \quad \|\text{diag}(H_k H_k^t)\| = \max_{j=1}^p (X_j X_k)^2 \leq \mu(X)^2,$$

and

$$\|\mathbb{E}M\| = \frac{s}{p} \|\text{diag}(HH^t)\|^2 = \frac{s}{p} \|H\|_{1 \rightarrow 2}^2 \leq \frac{s}{p} \|X\|^2.$$

Applying the NCCI completes the lemma. \square

Acknowledgment. The authors thank the referee for valuable comments that improved the paper. They thank Max Hügel for pointing out a mistake in a constant involved in a previous version of the tail-decoupling inequality.

REFERENCES

1. Bourgain, J., Tzafriri, L., Invertibility of "large" submatrices with applications to the geometry of Banach spaces and harmonic analysis. Israel J. Math. 57 (1987), no. 2, 137–224.
2. Candès, E. J. and Plan, Y. Near-ideal model selection by ℓ_1 minimization. Ann. Statist. 37 (2009), no. 5A, 2145–2177.
3. Ledoux, M. and Talagrand, M. Probability in Banach spaces. Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)], 23. Springer-Verlag, Berlin, 1991. xii+480 pp.
4. Oliveira, R. I. Concentration of the adjacency matrix and of the laplacian in random graphs with independent edges. arXiv:0911.0600, (2009).
5. de la Peña, V. H., and Montgomery-Smith, S.J. Bounds on the tail probability of U -statistics and quadratic forms. Bull. Amer. Math. Soc. (N.S.) 31 (1994), no. 2, 223–227.
6. de la Peña, V. H. and Giné, E. Decoupling. Probability and its Applications (New York). Springer-Verlag, New York, 1999.
7. Mendelson, S., Pajor A. and Tomczak-Jaegermann N. Uniform uncertainty principle for Bernoulli and sub-Gaussian ensembles, Constructive Approximation, 28 (2008), no. 3, 277–289.
8. Rudelson M. Random vectors in isotropic position, J. Funct. Anal. 164 (1999), no. 1, 60–72.
9. Rudelson, Mark; Vershynin, Roman. Sampling from large matrices: an approach through geometric functional analysis. J. ACM 54 (2007), no. 4, Art. 21, 19 pp.
10. Tropp, J. A. Norms of random submatrices and sparse approximation. C. R. Math. Acad. Sci. Paris 346 (2008), no. 23–24, 1271–1274.
11. Tropp, J. A. User friendly tail bounds for sums of random matrices. Found. Comput. Math. (2012), to appear. doi:10.1007/s10208-011-9099-z.

LABORATOIRE DE MATHÉMATIQUES, UMR 6623, UNIVERSITÉ DE FRANCHE-COMTÉ,
16 ROUTE DE GRAY,, 25030 BESANCON, FRANCE
E-mail address: `stephane.chretien@univ-fcomte.fr`

LATP, UMR 6632, UNIVERSITÉ DE PROVENCE, TECHNOPÔLE CHÂTEAU-GOMBERT,
39 RUE JOLIOT CURIE, 13453 MARSEILLE CEDEX 13, FRANCE
E-mail address: `darses@cmi.univ-mrs.fr`